Conclusions and questions from a non-randomised comparison of routine clinical services implementing different treatment models for borderline personality disorder

Barnicot K[1] and Crawford M[1]

[1]Department of Medicine, Imperial College London, Centre for Psychiatry, Commonwealth Building, Du Cane Road, London, W12 0NN, UK

Author for correspondence: Kirsten Barnicot, E-mail: k.barnicot@imperial.ac.uk

Financial support. The present study was funded by a Postdoctoral Research Fellowship awarded to KB from the United Kingdom National Institute for Health Research: PDF-2013-06-054.

Conflict of interest. None.

1495 words

Our study (Barnicot & Crawford 2018) was a non-randomised comparison of outcomes in routine clinical services implementing different treatment models for borderline personality disorder (BPD): mentalization-based therapy (MBT) and dialectical behaviour therapy (DBT). Our study's contribution lies primarily in its novelty as currently the only head-to-head comparison of these treatment models, and its potential to generate exploratory findings for further testing in a definitive randomised controlled trial. We thank Luyten and colleagues for their useful commentary.

Pre-treatment differences

Whilst pre-treatment dissociation and BPD severity were equivalent between treatment groups, DBT patients exhibited significantly greater emotional dysregulation and likelihood of PTSD, self-harming behaviour, and A&E admission in the 12 months before treatment (67% vs. 34%) (Barnicot & Crawford 2018, Table 1). Additionally, DBT patients exhibited clinically significantly, (although statistically non-significantly), higher rates of depression (43% vs. 25%), alcohol dependence (25% vs. 13%) and psychotic symptoms (46% vs. 35%). These may be chance differences, or viewed collectively, may raise questions about referral bias and participant self-selection, reinforcing the importance of randomised treatment allocation in future studies.

Reporting of findings

We agree that pre-registration and specification of primary outcomes is essential for randomised controlled trials. However, our study was not a clinical trial. We have fully reported and discussed all findings in our abstract, results and discussion.

Change in self-harm and emotional dysregulation

We based our conclusion that reductions in self-harm and emotional dysregulation were greater in DBT on our finding that, in both unadjusted *and* adjusted multi-level mixed-effects models, there was a significant treatment-by-time interaction favouring DBT for these outcomes. Two trials very recently published in this journal (July 2019 issue) used similar models to examine the rate of change in the outcome variables as their primary analysis (Freyer-Adam et al. 2019, Strauss et al. 2019), as did a largescale trial of MBT conducted by the treatment developer (Bateman et al. 2009). Further post-hoc analysis of our treatment-by-time interaction effect shows a larger decrease in emotional dysregulation and self-harm per month in DBT patients. MBT patients showed an estimated 1.08 point decrease in emotional dysregulation per month (β = -1.08, 95% CI -1.74 to –0.43, p = 0.001) compared to a 2.49 point estimated decrease per month in DBT patients (β = -2.49, 95% CI -3.15 to -1.83, p<0.001). MBT patients showed an estimated 0.95 times decrease in the rate of self-harm per month (IRR = 0.95, 95% CI 0.91 to 0.99, p =0.008) and DBT patients a 0.89 times estimated decrease per month (IRR = 0.89, 95% CI 0.86 to 0.92, p < 0.001). This fully justifies our conclusion that reductions in self-harm and emotional dysregulation were greater in DBT.

Luyten and colleagues usefully highlight the lack of between-groups difference in emotional dysregulation and self-harm at the 12-month follow-up (Barnicot & Crawford 2018, Table 2), suggesting therefore that patients followed different trajectories towards the same endpoint. Unlike our mixed-effects models – which benefited from incorporating data from multiple timepoints and maximum-likelihood-based modelling, increasing statistical power and robusticity to loss to follow-up (Tango 2016) - our sample size for outcome comparison at 12 months was restricted by using data from a single timepoint, and by loss to follow-up. Substantial overdispersion in self-harm incidents further reduced statistical power. We agree that the 6-point difference in emotional dysregulation score at month 12 is neither statistically nor clinically significant – thus whilst the rate of improvement in emotional dysregulation

was significantly faster amongst DBT patients, outcomes at the end of 12 months were similar. By contrast, there were large and clinically significant (but statistically non-significant) differences in self-harm. The median number of self-harm incidents between months 10 and 12 was 2 in the DBT condition and 12.5 in the MBT condition (Barnicot & Crawford 2018, Table 2 – medians are presented due to the skewed data distribution (Field 2013)). Thus, one possible interpretation is that the faster rate of improvement in self-harm in DBT was accompanied by a clinically significantly lower rate of self-harm by 12 months, which failed to reach statistical significance. A largescale adequately powered trial is required to determine whether there is a statistically significant difference in self-harm at 12 months or any other timepoint.

Reporting of findings from unadjusted and adjusted models

We strongly refute the suggestion that we have shown allegiance bias by only presenting adjusted models where findings from unadjusted models disadvantaged DBT. Firstly, whilst we are not privy to all unconscious biases that dwell within us, Barnicot is a non-clinical academic unaffiliated with any particular therapeutic paradigm; Crawford is a trained MBT therapist and delivered MBT in one of the participating treatment centres. Thus, any bias might be expected to favour MBT, not DBT. Secondly, our choice of when to present adjusted models was based solely on an objective criterion: if an outcome was found to differ significantly between DBT and MBT patients in an *unadjusted model*, we presented findings from an *adjusted* model, in order to test whether the significant difference in the unadjusted model still held after adjusting for potential confounders. If an outcome was *not* found to differ significantly between DBT and MBT patients in *unadjusted* models, we did not run an adjusted model (Barnicot & Crawford 2018, Methods). This criterion applied both to findings favouring DBT and those favouring MBT.

Treatment dropout

Noting the markedly higher treatment dropout rate in DBT, and cautious of attributing any outcome differences to inherent effects of therapy modality that might instead be better explained by treatment dropout or factors linked to treatment dropout, we conservatively chose to adjust our models for any such factors. Therefore, as well as including baseline covariates differing significantly between DBT and MBT patients, we also included baseline covariates differing significantly between treatment completers and dropouts (Barnicot & Crawford 2018, Methods). Guidance on clinical trial analysis agrees that adjustment for treatment dropout can be useful in sensitivity analyses (European Medicines Agency 2015). However, we agree with Luyten and colleagues that caution is merited, as if a higher treatment dropout rate is a true property of DBT treatment and also associated with poorer outcomes, then adjusting for variance shared between DBT treatment and treatment dropout could obscure evidence of poorer outcomes in DBT (European Medicines Agency 2015). Our analyses were conducted on an intention-to-treat basis, including data collected from treatment dropouts following their treatment discontinuation. The requested replication of our analyses in the sample of treatment completers alone would restrict our sample size to n = 43 at baseline and n = 40 at month 12, reducing the power of the study to detect any treatment differences. We do not believe this would be useful.

Covariate-adjusted models in non-randomised studies

Luyten and colleagues cite an important difficulty with adjustment for covariates whose effects overlap with treatment effects. This is particularly problematic in non-randomised studies, where it is impossible to determine whether pre-treatment differences between groups have arisen by chance or whether they constitute an inherent property of group membership, and thus by removing variance that is shared between the covariate and

treatment exposure, adjustment for group differences may obscure or enhance treatment effects (Field 2013, Miller & Chapman 2001). However, this problem arises when covariates are related to *treatment exposure*, not to the dependent variable (Field 2013, pg. 484). Thus, the request that we exclude any covariates associated with our dependent variables would not solve the problem of shared variance with the treatment effect and would be contrary to clinical trial guidance (European Medicines Agency 2015). Despite this difficulty, UK National Institute for Health and Care Excellence (NICE) Decision Support Unit guidance stresses the importance of evaluating and adjusting for potentially confounding differences between treatment groups in non-randomised studies (Faria et al. 2015). A possible compromise may be found in the NICE recommendation that authors present sensitivity analyses to test whether their findings are robust to different analysis methods (Bell et al. 2016).

Conclusion

Therefore, acknowledging the inherent difficulty of covariate adjustment given our non-randomised study design, and yet also cognisant of the recommendation that consideration of confounders is important in such studies, we propose that our covariate-adjusted models be considered sensitivity analyses and their findings exploratory. Our unadjusted and adjusted analyses agree that the rate of change in emotional dysregulation and self-harm was superior amongst DBT patients. Thus, these findings seem robust to different analysis methods. Our unadjusted analyses also found statistically and clinically significantly higher rates of treatment dropout, A&E attendance and psychiatric hospitalisation amongst DBT patients, which were not apparent in adjusted models. This evidence of poorer outcomes amongst DBT patients merits further investigation. Whilst our adjusted analyses suggested these outcomes may have been influenced by differences between DBT and MBT patients in pre-treatment characteristics and treatment dropout, this does not detract from the importance of these

negative outcomes, and the difficulties of adjusting for treatment dropout and interpreting covariate-adjusted analyses in non-randomised studies should be held in mind. Further evaluation using a randomised controlled trial design is essential in order to draw robust conclusions on these and other outcomes.

Barnicot K and Crawford M (2018). Dialectical behaviour therapy v. mentalisation-based therapy for borderline personality disorder. *Psychological Medicine* 1–9. https://doi.org/10.1017/ S0033291718002878

Bateman A and Fonagy P (2009) Randomized controlled trial of outpatient mentalization based treatment versus structured clinical management for borderline personality disorder. *American Journal of Psychiatry* 166, 1355–1364.

Bell H, Wailoo AJ , Hernandez M , Grieve R , Faria R, Gibson L and Grimm S (2016). The use of real world data for the estimation of treatment effects in NICE decision making. Report by the Decision Support Unit, National Institute for Health and Care Excellence, United Kingdom. (http://nicedsu.org.uk/wp-content/uploads/2018/05/RWD-DSU-REPORT-Updated-DECEMBER-2016.pdf) Accessed 12th August 2019

European Medicines Agency Committee for Medicinal Products for Human Use (2015). Guideline on adjustment for baseline covariates in clinical trials. (https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf.). Accessed 12th August 2019

Faria R, Hernandez Alava M, Manca A and Wailoo AJ (2016). The use of observational data to inform estimates of treatment effectiveness in technology appraisal: Methods for comparative individual patient data. Report by the Decision Support Unit, National Institute for Health and Care Excellence, United Kingdom. (http://nicedsu.org.uk/wp-

content/uploads/2016/03/TSD17-DSU-Observational-data-FINAL.pdf) Accessed 12th August 2019

Field A (2013) *Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock 'n' roll, 4th edition.* Sage: London.

Freyer-Adam J, Baumann S, Haberecht K, Bischof G, Meyer C, Rumpf H-J, John U and Gaertner B (2019). Can brief alcohol interventions in general hospital inpatients improve mental and general health over 2 years? Results from a randomized controlled trial. *Psychological Medicine* 49, 1722–1730. https://doi.org/10.1017/S0033291718002453

Miller GA, Chapman JP (2001). Misunderstanding Analysis of Covariance. *Journal of Abnormal Psychology* 110(1), 40-48.

Strauss GP, Granholm E, Holden JL, Ruiz I, Gold JM, Kelly DL and Buchanan RW (2019). The effects of combined oxytocin and cognitive behavioral social skills training on social cognition in schizophrenia. *Psychological Medicine* 49, 1731–1739. https://doi.org/10.1017/S0033291718002465

Tango T (2016). On the repeated measures designs and sample sizes for randomized controlled trials. *Biostatistics* 17(2), 334–349.